Review of:
The All Convolutional Net

SiKai Lee
Columbia University
116th St & Broadway, New York, NY, USA
sikai.lee@columbia.edu

ZhiHao Luo
Columbia University
116th St & Broadway, New York, NY, USA
zl2463@columbia.edu

Juan Pablo Colomer
Columbia University
116th St & Broadway, New York, NY, USA
jpc2192@columbia.edu

Abstract

In this report, we review the paper The All Convolutional Net by Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller.
Instead of building convolutional neural nets with the conventional principle, the authors experiment with models that are increasingly homogeneous. They find that both the max-pooling layer and the fullying connected layer – which have been important parts of the CNN structures – can be replaced by other convolutional layers. This new structure with almost purely convolutional layer is said to perform as well as the state-of-the-art CNN architectures if not better in some cases.
***TODO:** We find that...*

1. Review

1.1. Convolutional Neural Nets Background

In recent years, Convolutional Neural Nets (CNNs) have become more and more popular for image related tasks. First introduced by Yann Lecun, this particularly powerful neural network structure has become wide adopted by researches for their efficiency and robustness. Over the years, multiple papers attempts to improve performance of the CNN by changing activation functions or by augmenting the structure. However, most designs seem to share the same principle: they use alternating convolution and max-pooling layers followed by some number of fully connected layers. A typical convolution maxpooling sequence is shown in figure 1.

The convolution layers defines a set of filters and an activation function. It convolve the input to that layer with its filters to generate an output which is then passed through the activation function. In this example, the convolution layer contains $2F$ filters of size $P \times P$; by convolving the $N \times N$ input with all of its filters, it generate $2F$ feature maps with size $(N - P + 1) \times (N - P + 1)$. The final output of the convolution layers are obtained by passing the feature maps through the ReLU function.

$$ReLU(x) = max(0, x) \tag{1}$$

The max-pooling layer down-scales the feature maps by dividing the feature maps into subsections and representing that entire subsection with the maximum value inside it.
Finally, at the end of the last max-pool layer, the feature maps are flattened into a vector which are usually used as the input to the fully connected layers are shown in Figure 2
In the fully connected layers, each node has a weighted con-

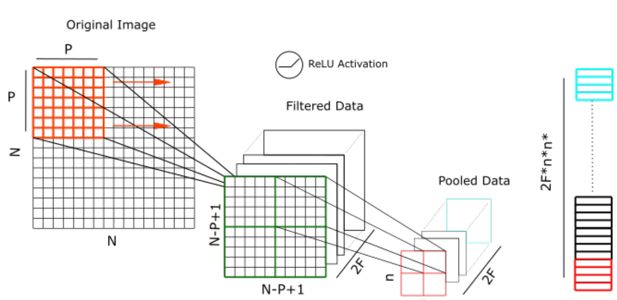

Figure 1. An example architecture of a CNN. On the left is an 2D image input, followed by a convolutional layers, and a max pooling layer.

nection to every node from the previous layer, it's output is determined by passing through the weighted sum of previous layer output through an activation function. The input is non-linear transformed through the layers and this non-linear transformation helps the neural net in finding separation between classes.

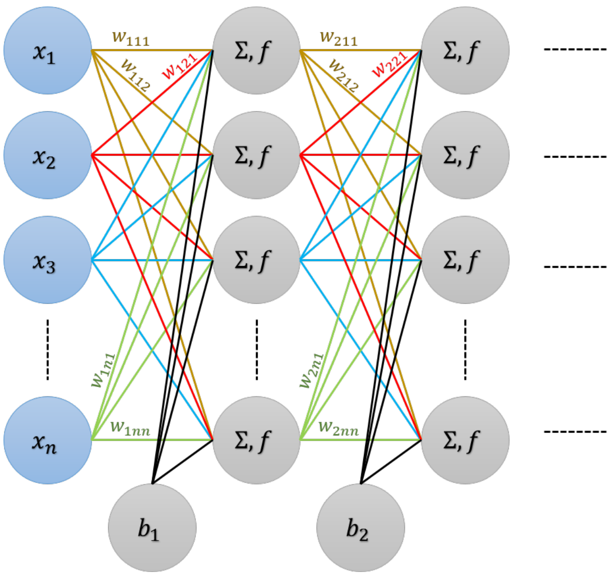

Figure 2. Fully connected layers

1.2. Homogeneous Structure

The author challenges the conventional CNN structure by introducing a number of models that are increasing more homogeneous – the most extreme of which contains only convolutional layers without pooling or fully connected layers.

2. Experiments

some words

3. Discussion

some words